3.0 *IN VITRO* HUMAN SKIN MODEL SYSTEMS FOR SKIN CORROSION

3.1 Background

Pre-validation and validation studies have been completed for an *in vitro* human skin model system commercially available as EPISKIN™ (ICCVAM 2002; Fentem et al. 1998; Botham et al. 1992; Botham et al. 1995; Barratt et al. 1998). Based on its scientific validity, this test method has been recommended for the testing of all classes of chemicals (ICCVAM 2002; Fentem et al. 1998; Balls and Corcelle 1998b) and for inclusion in tiered testing strategies as part of a tiered or weight-of-evidence evaluation (ICCVAM 2002). In addition to EPISKIN™, a related human skin model corrosivity test method marketed as EpiDerm™ (EPI-200) has been validated (Liebsch et al. 2000). Neither test method has been validated for categorizing the corrosive properties of chemicals across the three UN Packing Group subcategories of corrosivity (ICCVAM 2002; Liebsch et al. 2000; Balls and Hellsten 2000). This chapter briefly describes the principles of *in vitro* human skin model systems for corrosivity followed by the recommended performance standards, which consists of essential test method components, reference chemicals, and comparison of accuracy and reliability.

3.2 Principles of *In Vitro* Human Skin Model Systems for Skin Corrosion

The test material is applied topically to a three-dimensional human skin model, comprised of at least a reconstructed epidermis with a functional stratum corneum. Corrosive materials are identified by their ability to induce a decrease in cell viability below defined threshold levels at specified exposure periods. The principle of the human skin model assay is based on the premise that corrosive chemicals are able to penetrate the stratum corneum by diffusion or erosion, and are cytotoxic to the cells in the underlying layers. The use of test systems that include human-derived cells or tissue should be in accordance with applicable national and international laws, regulations, and policies.

Investigators using an *in vitro* human skin model system for skin corrosion must be able to demonstrate that the assay is valid for its intended use. This includes demonstrating that different preparations are consistent in barrier properties (i.e., capable of maintaining a barrier to noncorrosive substances, able to respond appropriately to weak and strong corrosive substances) and/or that any modification to the existing validated reference test method does not adversely affect its performance characteristics.

In vitro human skin model systems for skin corrosion may be used to test solids, liquids, and emulsions of any chemical or product class. The liquids can be aqueous or nonaqueous; solids can be soluble or insoluble in water. The samples may be pure chemicals, dilutions, formulations, or waste. Where appropriate, solids should be ground to a powder before application; no other prior treatment of the sample is required. In some chemical classes, relatively few chemicals were included in the validation of the accepted *in vitro* human skin model system for skin corrosion (Fentem et al. 1998). However, taking into account the limited mechanisms that result in corrosivity, this method is expected to be generally applicable across all chemical classes (ICCVAM 2002; Fentem et al. 1998; Balls and Corcelle 1998b).

3.3 Essential Test Method Components

The following is a description of the essential test method components for *in vitro* human skin model test methods for skin corrosivity, as provided in OECD Test Guideline 431 (OECD 2003a). Human skin models can be obtained commercially (e.g., EPISKIN™, EpiDerm™ [EPI-200]) or they can be developed or constructed in the testing laboratory (Ponec et al. 2000; Wilkins et al. 1994).

3.3.1 *In Vitro* Human Skin Model Conditions

Human keratinocytes should be used to construct the epithelium. Multiple layers of viable epithelial cells should be present under a functional stratum corneum. The skin model may also have a stromal component layer. Stratum corneum should be multilayered with the necessary lipid profile to produce a functional barrier with robustness to resist the rapid penetration of cytotoxic chemicals used as positive controls. The containment properties of the model should prevent the passage of material around the stratum corneum to the viable tissue, which would lead to poor modeling of the exposure to skin. The skin model should be free of contamination with bacteria, mycoplasma, or fungi.

The magnitude of viability is usually quantified by using MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, thiazolyl blue; CASRN 298-93-1) or other metabolically converted vital dyes (Marshall et al. 1995). The negative control tissue should be stable in culture (provide similar viability measurements) for the duration of the test exposure period. The stratum corneum should be sufficiently robust to resist the rapid penetration of positive control chemicals (e.g., 1% Triton X-100), which can be assessed by the exposure time required to reduce cell viability by 50%.

3.3.2 Application of the Test Substances

Two tissue replicates are used for each test and control substance. For liquid materials, sufficient test substance must be applied to uniformly cover the skin surface; a minimum of 25 $\mu L/cm^2$ should be used. For solid materials, sufficient test substance must be applied evenly to cover the skin surface, and it should be moistened with deionized or distilled water to ensure good contact with the skin. Where appropriate, solids should be ground to a powder before application. At the end of each exposure period (3 minutes to 1 or 4 hours), the test material must be carefully washed from the skin surface with an appropriate buffer or 0.9% NaCl.

3.3.3 Control Substances

Solvent Controls: In tests that involve the use of a vehicle or solvent with the test substance, the vehicle or solvent must be compatible with the barrier system (i.e., not alter the integrity of the membrane barrier system) and should not alter the corrosivity of the test substance. When applicable, solvent (or vehicle) controls should be tested concurrently with the test substance to demonstrate the compatibility of the solvent with the barrier system.

Positive (Corrosive) Controls: A positive control chemical should be tested concurrently with the test substance to demonstrate that the human skin membrane barrier is functioning properly. The positive control should be well characterized for its corrosive activity and should generate a response that is low to intermediate within the range of corrosive responses for the assay. An acceptable

positive control response range must be developed based on historical positive control(s) data. In each test, the positive control should be evaluated to determine if the value is within the acceptable positive control range. Typically, for biologically-based test methods, acceptable ranges are within 2 to 3 standard deviations of the historical mean response but developer of proprietary test methods may establish tighter ranges. Glacial acetic acid is an example of a positive control substance producing a low to intermediate response in the validated reference test method.

Negative (Noncorrosive) Controls: A noncorrosive substance should also be tested concurrently with the test substance as another quality control measure to demonstrate the functional integrity of the human skin membrane barrier. Examples of noncorrosive substances used as negative controls in the validated reference test method include 0.9% sodium chloride and water.

Benchmark Controls: Benchmark controls may be useful to demonstrate that the test method is functioning properly for detecting the dermal corrosivity potential of chemicals of a specific chemical class or a specific range of responses, or for evaluating the relative corrosivity potential of a corrosive test substance. Appropriate benchmark controls should have the following properties:

- consistent and reliable source(s) for the chemical
- structural and functional similarity to the class of the substance being tested
- known physical/chemical characteristics
- supporting data on known effects in animal models
- known potency in the range of response (including moderate response)

3.3.4 Viability Measurements

Only standardized, quantitative methods should be used to measure cell viability. Furthermore, the measure of viability must be compatible with use in a three-dimensional tissue construct. Non-specific dye binding must not interfere with the viability measurement. Protein binding dyes and those that do not undergo metabolic conversion (e.g., neutral red) are therefore not appropriate. The most frequently used assay is MTT reduction, which has been shown to give accurate and reproducible results (Fentem et al. 1998) but others may be used.

Chemical action by the test material on the vital dye may mimic that of cellular metabolism leading to a false estimate of viability. This has been shown to happen when such a test material is not completely removed from the reconstructed skin by rinsing (Liebsch et al. 2000). If the test material directly acts on the vital dye, additional controls should be used to detect and correct for test substance interference with the viability measurement (Liebsch et al. 2000; Fentem et al. 2001).

3.3.5 Interpretation of Results

The optical density (OD) values obtained for each test sample can be used to calculate percentage viability relative to the negative control, which is arbitrarily set at 100%. The cell viability criteria used to distinguish between corrosive and noncorrosive test chemicals (or to discriminate between different corrosive classes), or the statistical procedure(s) used to evaluate the results and identify corrosive materials must be clearly defined and documented, and be shown to be appropriate. In general, such criteria are established during test optimization, tested during a prevalidation phase,

and confirmed in a validation study. As examples, the predictions of corrosivity associated with EPISKIN™ (Fentem et al. 1998) and EpiDerm™ (EPI-200) (Liebsch et al. 2000) are:

EPISKIN™: The test substance is considered to be corrosive to skin:

i) if the viability after 3 minutes of exposure is less than 35%, or
ii) if the viability after 3 minutes of exposure is greater than or equal to 35% and the viability after 4 hour of exposure is less than 35%.

The test substance is considered to be noncorrosive to skin:

i) if the viability after 4 hours of exposure is greater than or equal to 35%.

EpiDerm™ (EPI-200): The test substance is considered to be corrosive to skin:

i) if the viability after 3 minutes of exposure is less than 50%, or
ii) if the viability after 3 minutes of exposure is greater than or equal to 50% and the viability after 1 hour of exposure is less than 15%.

The test substance is considered to be noncorrosive to skin:

i) if the viability after 3 minutes of exposure is greater than or equal to 50% and the viability after 1 hour of exposure is greater than or equal to 15%.

3.3.6 Test Report

The test report should include the following information, if relevant to the conduct of the study:

Test and Control Substances

- Chemical name(s) such as CAS preferred name and RN, followed by other names, if known
- Purity and composition of the substance or preparation (in percentage(s) by weight)
- Physicochemical properties such as physical state, volatility, pH, stability, chemical class, water solubility relevant to the conduct of the study
- Treatment of the test/control substances prior to testing, if applicable (e.g., warming, grinding)
- Stability, if known

Justification of the Skin Model and Protocol Used

Test Method Integrity

- The procedure used to ensure the integrity (i.e., accuracy and reliability) of the test method over time
- If the test method employs proprietary components, the procedure used to ensure their integrity from "lot-to-lot" and over time
- The procedures that the user may employ to verify the integrity of the proprietary components

Criteria for an Acceptable Test

- Acceptable concurrent negative control ranges based on historical data
- Acceptable concurrent positive control ranges based on historical data

Test Conditions

- Cell system used
- Calibration information for measuring device used for measuring cell viability (e.g., spectrophotometer)

- Complete supporting information for the specific skin model used including its validity
- Details of test procedure used
- Test doses used
- Description of any modifications of the test procedure
- Reference to historical data of the model
- Description of evaluation criteria used

Results

- Tabulation of data from individual test samples (e.g., OD values and calculated percentage cell viability data for the test substance and the positive, negative, and benchmark controls, reported in tabular form, including data from replicate repeat experiments as appropriate, and means and ± the standard deviation for each trial)

Description of Other Effects Observed
Discussion of the Results
Conclusion

3.4 Reference Chemicals

To demonstrate technical proficiency with the validated reference test method, the user should evaluate his/her ability to correctly identify the dermal corrosivity classification of twelve of the chemicals (6 noncorrosive and 6 corrosives varying in corrosive potency) listed in Table 3-1. However, to ensure that a proposed *in vitro* skin model system possesses reliability and accuracy characteristics that are comparable to the validated reference test method, the 24 reference chemicals listed in **Table 3-1** must be used. The 24 reference chemicals (12 noncorrosives, 12 corrosives) listed in **Table 3-1** provide a representative distribution of the 60 chemicals used in the ECVAM validation study of EPISKIN™ (Fentem et al. 1998; Barratt et al. 1998) and cover the range of corrosivity responses obtained for the *in vivo* rabbit skin reference test method. The 24 reference chemicals include 23 of the 24 chemicals used to validate EPIDERM™ (EPI-200), a test method structurally and functionally similar to EPISKIN™ (Liebsch et al. 2000). Included in this list are five organic bases, four inorganic acids, three inorganic bases, three organic acids, three electrophiles, three phenols, two neutral organics, and one surfactant. These reference chemicals are the minimum number that should be used to evaluate the performance of a mechanistically and functionally similar, proposed test method. These chemicals should not be used to develop the prediction model for a proposed test method. If any of the recommended chemicals are unavailable, other chemicals for which adequate reference data are available could be substituted. To the extent possible, the substituted chemical(s) should be of the same chemical class as the original chemical(s). If desired, additional chemicals representing other chemical or product classes and for which adequate reference data are available can be used to more comprehensively evaluate the accuracy of a proposed test method. However, these additional chemicals should not include any that had been used to develop the prediction model for the proposed test method.

3.5 Accuracy and Reliability

When evaluated using the minimum list of recommended reference chemicals (**Table 3-1**), the proposed test method should have reliability and performance (i.e., sensitivity, specificity, false positive rates, and false negative rates) characteristics that are comparable to the performance of the validated reference test method (ICCVAM 2002; Fentem et al. 1998). Noncorrosive and corrosive

chemicals, ranging in activity from strong to weak, and representing relevant chemical classes are included so that the performance of the proposed test method can be determined and compared to that of the validated reference test method. Eleven of the 12 chemicals mentioned in the OECD proposed Test Guideline 431 (*In vitro* skin corrosion human skin model system) (OECD 2003a) are included. Acrylic acid, proposed by the OECD as a severe corrosive, was not included because the comparative performance of this chemical in the validated reference test method (EPISKIN™) and the *in vivo* rabbit skin corrosivity test had not been demonstrated and thus the accuracy of the validated reference test method for this chemical was not established. Based on experience with the validation of different *in vitro* test methods, one effective approach used to establish intra- and inter-laboratory reproducibility for a test method not previously validated is to test each of the reference chemicals three times in each of three independent laboratories. The accuracy of the validated *in vitro* human skin model test system, EPISKIN™, for the 24 reference chemicals and the complete validation database considered by ICCVAM are provided in **Table 3-2**. Its accuracy for the reference chemicals and the corresponding values obtained for the total database compiled during the ICCVAM evaluation process are not identical due to constraints associated with the chemical selection process.

Table 3-1 Recommended Chemicals for Validation of New *In Vitro* Human Skin Model Corrosivity Test Methods

Chemical[1]	CASRN	Chemical Class[2]	UN *In Vivo* PG	pH[3]
***In Vivo* Corrosives**				
Phosphorus tribromide	7789-60-8	inorganic acid	I	1.0
Sulfuric acid (10%)	7664-93-9	inorganic acid	II/III	1.2
Boron trifluoride dihydrate	13319-75-0	inorganic acid	I	1.5
Glycol bromoacetate (85%)	3785-34-0	electrophile	II/III	2.0
Caprylic acid	124-07-02	organic acid	II/III	3.6
2-tert-Butylphenol	88-18-6	phenol	II/III	3.9
Dimethyldipropylenetriamine	10563-29-8	organic base	I	8.3
Dimethylisopropylamine	996-35-0	organic base	II/III	8.3
1,2-Diaminopropane	78-90-0	organic base	I	8.3
n-Heptylamine	111-68-2	organic base	II/III	8.4
2-Mercapoethanol, sodium salt (45% aq.)	37482-11-4	inorganic base	II/III	12.0
Potassium hydroxide (10% aq.)	1310-58-3	inorganic base	II	13.1
***In Vivo* Noncorrosives**				
Sulfamic acid	5329-14-6	inorganic acid	NC	1.5
Isostearic acid	30399-84-9	organic acid	NC	3.6
Phenethyl bromide	103-63-9	electrophile	NC	3.6
Eugenol	97-53-0	phenol	NC	3.7
1,9-Decadiene	1647-16-1	neutral organic	NC	3.9
o-Methoxyphenol	90-05-1	phenol	NC	3.9
Sodium lauryl sulfate (20% aq.)	151-21-3	surfactant	NC	3.9

Chemical[1]	CASRN	Chemical Class[2]	UN *In Vivo* PG	pH[3]
Tetrachloroethylene	127-18-4	neutral organic	NC	4.5
4-Amino-1,2,4-triazole	584-13-4	organic base	NC	5.5
4-(methylthio)-Benzaldehyde	3446-89-7	electrophile	NC	6.8
Sodium carbonate (50% aq.)	7664-93-9	inorganic base	NC	11.7
Dodecanoic acid (lauric acid)	143-07-7	organic acid	NC	ND

Abbreviations: aq = aqueous; CASRN = Chemical Abstracts Service Registry Number; PG = Packing Group; NC = Noncorrosive; ND = not determined (unable to measure); UN = United Nations.

[1]These chemicals, sorted first by corrosives versus noncorrosives and then by pH, were selected from among the 60 chemicals used by ECVAM to validate EPISKIN™ (Fentem et al. 1998; Barratt et al. 1998). Unless otherwise indicated, the chemicals were tested at the purity level obtained when purchased from a commercial source (Barratt et al. 1998). The goal of the selection process was to include, to the extent possible, chemicals that: were representative of the range of corrosivity responses (e.g., noncorrosives; weak to strong corrosives) that the validated reference test method is capable of measuring or predicting; were representative of the chemical classes used in the validation process; reflected the performance characteristics of the validated reference test method; have chemical structures that were well-defined; induced reproducible results in the validated reference test method; induced definitive results in the *in vivo* reference test; were commercially available; and were not associated with prohibitive disposal costs.

[2]Chemical class assigned by Barratt et al. (1998).

[3]The pH values were obtained from Fentem et al. (1998) and Barratt et al. (1998).

The reliability of the proposed test method for the reference chemicals should be comparable to that of the validated reference test method. An assessment of interlaboratory reproducibility is not essential if the test method is to be used in one laboratory only. In terms of cell viability measurements, the median CV should not exceed 35% for studies conducted in different laboratories (ICCVAM 2002; Fentem et al. 1998). The median CV for replicate studies conducted in the same laboratory should be appreciably less than the median CV for studies conducted in different laboratories.

Table 3-2 Accuracy of the Validated *In Vitro* Human Skin Model System Test Method (EPISKIN™) for Skin Corrosion[1]

Source	# of Chemicals	# of Tests[2]	Sensitivity	Specificity	False Negative Rate	False Positive Rate
Reference Chemicals	24	216	83% (90/108)	79% (85/108)	17% (18/108)	21% (23/108)
Fentem et al. (1998)	60	540	83% (201/243)	80% (237/297)	17% (42/243)	20% (60/297)

Definitions: Sensitivity is defined as the proportion of all positive chemicals that are correctly classified as positive in a test. Specificity is defined as the proportion of all negative chemicals that are correctly classified as negative in a test. False positive rate is defined as the proportion of all negative chemicals or chemical mixtures that are falsely identified as positive. False negative rate is defined as the proportion of all positive chemicals or chemical mixtures that are falsely identified as negative.

[1]The ability of the validated *in vitro* human skin model system to correctly predict the *in vivo* rabbit skin corrosivity potential of the 24 reference chemicals and the corresponding performance characteristics obtained by Fentem et al. (1998) are not identical due to the constraints associated with selection of the reference chemicals. The goal of the selection process was to include, to the extent possible, chemicals that: were representative of the range of corrosivity responses (e.g., noncorrosives; weak to strong corrosives) that the validated reference test method is

capable of measuring or predicting; were representative of the chemical classes used during the validation process; reflected the performance characteristics of the validated reference test method; have a chemical structure that was well-defined; induced reproducible results in the validated reference test method; induced definitive results in the *in vivo* reference test; were commercially available; and were not associated with prohibitive disposal costs.

[2]In the Fentem et al (1998) validation study, each chemical was tested three times in each of three laboratories. Due to the presence of a balanced design, the performance characteristics are based on individual tests rather than individual chemicals.